

UEFA Champions League Data Analysis Project

Alex Eidt and Pranav Natarajan

Research Questions and Methodology

The UEFA CHAMPIONS LEAGUE DATASET SUBSET from 2008-2015

Based on past performances in the Champions League, which team is most likely to win the competition?

Compare teams based on their record against other teams.

1. What are the win percentages of all teams in the competition by year?
 - a. Visualize Win percentages of Top 10 teams according to win percentage against teams with varying win percentages
2. What are the average goals scored and aG (Actual goals) for a team in a given year by win percentage?
3. What are the average goals conceded and the number of goals conceded in a game for a team by year?
4. Compile above data by year into a DataFrame
5. We now can use our model to get the coefficient! Coefficients are weighted the stage of the competition the game took place. (Later stages = higher weight).
 - . Get the coefficient data, Visualize the win percentages of team by year and the coefficient changes over the years for the teams

Perform Machine Learning using a regressor() model to get the expected goals by a team, and therefore direct comparisons to other teams using the coefficient to declare a winner, comparing how our model fared with actual results of the winners over the years.

Dataset

The Dataset can be found at this URL:

<https://github.com/jalapic/engsoccerdata/blob/master/data-raw/champs.csv>

This is a dataset that contains data on the scores of UEFA Champions League Matches from 1955 to 2017. The data describes the entire champions league competition for that season, indicating which level of the competition (semi-finals, quarter-finals, etc) the game took place in. We will subset this data to exclude qualifier match data and contain data beginning from the group stages of the competition for each of those years.